

ABSTRACT OF THE DISCLOSURE

5 An audio processing system and method for providing synchronized display of recognized text from an original audio file and playback of the original audio file. The system includes a speech recognition module, a silence insertion module, and a silence detection module. The speech recognition module generates text and audio pieces. The silence insertion
10 module aggregates the audio pieces into an aggregated audio file. The silence detection module converts the original audio file and the aggregated audio file into silence detected versions. Silent and non-silent blocks are identified using a threshold volume. The silence insertion module compares the silence detected original and aggregated audio files, determines the
15 differences in position of non-silence elements and inserts silence within the audio pieces accordingly. The characteristics of the silence inserted audio pieces are used to synchronize the display of recognized text from an original audio file and playback of original audio file.